从朗道势垒到奥数 -- 我对未来教育和科学的看法

作者 顾立帆

—— 水往低处流。水可能还会往前流一点，但失去了部分势能，还有各种耗散。

Turing Test 和主流教科书 AI: A Modern Approach 的哲学基础和测试理论问题，及权威学术杂志 Nature 上 AlphaGo Zero 论文中的错误，显示了当前教育和科学研究的危机 [1]。

1949 年诺贝尔生理和医学奖 leukotomy 就存在严重的实验检测问题，造成灾难性误导。在科学研究上，有错不纠，会使危机越演越烈。

这些问题显示：对科学的基础和方法，有严重误解。科学起源于哲学。科学的基础，不是数学，不是经验，而是证伪实验。科学实验也不同于经验测试。要理解人工智能和计算机的局限性，研究人和人工智能更好合作模式，需要建立更准确的智力概念和语言。

过分注重技巧，排斥基础理论，排斥对科学的哲学基础的研究和分析，也会纵容学术上的诡辩盛行，瓦解科学的基础，危及未来的发展。

本文从 leukotomy 开始，分析教育和科学研究面临的危机和挑战。

一、Leukotomy 和 BRIAN Initiative

Leukotomy 造成严重问题，原因包括：

1）选择手术对象时没有区分 personalities 和 mental diseases。由于对 personalities 缺乏研究，目前仍无法准确区分两者。这也使得使用基因技术的 BRAIN Initiative 项目更加危险。

2）给 leukotomy 授与诺贝尔奖时，没有正确地检验 leukotomy 的后果，对提出警告者也没有重视。而本人对 BRAIN Initiative 提出的警告和对 AlphaGo 系列提出的检验要求，也没有被重视。

3）该手术对 personality 和 intelligence 有严重损伤。连大规模实施者 Walter Freeman 都在一定程度上承认。由于缺乏对 personality 和 intelligence 的理解，事后的统计并不正确 [2]。

而无法理解 personality 和 intelligence，是由于缺乏正确描述 personality 和 intelligence 的语言、缺乏开拓科学新学科所需要的哲学性思考和分析，而造成的无知。

另外，leukotomy 获得诺贝尔奖之前的研究，没有作严格控制条件的科学实验，仅仅凭经验性的手术尝试作出主观的评估。这种主观评估跟后来发现的灾难性后果明显不符。

而缺乏严格控制实验条件的主观评估，也正是 Turing Test 的问题所在。Nature 上 AlphaGo Zero 论文中的 superhuman 宣称，也是缺乏严格控制条件的实验验证的主观评估。因为 superhuman 是针对 generic 的 human，只要有人能设计出证伪的实验，就可以否定这宣称。

对 leukotomy 的诺贝尔奖，有错不纠，和 Nature 上 AlphaGo Zero 论文中的 superhuman 宣称拒绝严格控制条件的实验

证伪，显示 peer review 有严重问题。Peer review 也是主观评估。

从 leukotomy 开始的问题已经危及脑和智力研究及相关领域，包括 BRAIN Initiative 项目。

二、Technological Singularity 和 AlphaGo 系列的问题

Nature 杂志上的 AlphaGo Zero 论文，宣称 AlphaGo Zero 取得了 superhuman 的 performance。但 superhuman 的概念跟 generic 的 human 有关。该论文并没有提供任何 superhuman 的证据。

AlphaGo Zero 击败 AlphaGo Master 并不是 superhuman 的证据，AlphaGo Master 在一些对弈中击败一些人类棋手，也不是 generic 的 superhuman 的证据。这种 game 只是经验测试，不是严格控制条件的科学实验。

只要有人设计出能证明 AlphaGo Zero 不是 superhuman 的证伪实验，就可以否定该论文中的这一宣称。但 AlphaGo Zero 至今拒绝本人设计的证伪实验。

科学的基础是证伪，不是经验。逻辑实证主义在科学哲学上是有问题的。因为逻辑判断使用的语言和经验证据之间可能有很大的差别，也就是说经验证据往往不能真正支持逻辑判断。

比如，不管地球上有多少人看到太阳东升西落，也不能证明太阳回绕地球转。不管电脑围棋多少次在 game 中战胜人类，也不能证明电脑围棋已经 superhuman 了。

再多的经验证据，也不能得出科学结论。只要有人能设计出严格控制实验条件的证伪实验，就可以否定经验结论。

近年有人根据人工智能的一些成功经验，炒作 Technological Singularity，实际上是对科学哲学的无知。会误导了脑和智力的研究。

我无意贬低电脑围棋的进展。事实上，我是鼓励电脑围棋的研究。早在 Deepmind 成立之前，我就分析认为电脑围棋可以击败人类顶级棋手，因为围棋是在很小空间里的高度规则化游戏 [3]。这已被证实。但这并不意味电脑围棋有通用智力潜力或超人能力。

AlphaGo Master 击败柯洁后，很多人开始膜拜 AlphaGo Master。而我指出：AlphaGo Master 虽然赢棋，但只要不是最优解，就可能走的是无理棋。这也已被证实，因为 AlphaGo Zero 击败 AlphaGo Master 意味着 AlphaGo Master 很多离谱的走法其实是无理棋。

见到的 AlphaGo Zero 棋谱中，对 AlphaGo Master 的二十局是 AlphaGo Zero 最后版（或稳定版）[4]，有很不寻常现象：AlphaGo Master 执黑的十盘棋中，布局至 16 手完全一样。而 AlphaGo Zero 执黑的十盘棋中，布局至 18 手完全一样，AlphaGo Zero 输棋的 14、16 局，至 45 手完全一样。所以有如下问题：

1）在 Nature Magazime 上发表两篇 AlphaGo 系列文章后，Deepmind 有没有在这二十局棋里把 AlphaGo Zero 和 AlphaGo Master 对局的情况，全面准确地告诉公众？如果答案是，则：

2) 公布的这二十局棋谱中，所出现的布局高度类同，是由于 AlphaGo Zero 接近围棋最优解，还是陷入 local trap？如果认为是前者，需要用严格控制条件的严谨实验来验证。如果是后者，AlphaGo Zero 走的也是无理棋，而 superhuman 的宣称可能对围棋和智力科学研究带来严重误导。

其实 AlphaGo Zero 击败 AlphaGo Master，本身就是人类击败 AlphaGo Master 的证据：人类靠设计新算法系统击败了 AlphaGo Master，而 AlphaGo Master 自身的进化速度远不及人类根据系统结构分析来设计新算法的速度 [5]。Nature 杂志上的 AlphaGo Zero 论文发表前几个月，我就针对 AlphaGo Master 怀疑：利用自我对弈收集数据和 deep-learning 来涨棋，几年后其效果会不会暴露出局限？事实上，AlphaGo Master 的弱点不到四个月就暴露。

即使人类能设计更新的算法击败 AlphaGo Zero [6]，仍不能证明电脑围棋是 superhuman。我已经设计了针对 AlphaGo Zero 的 superhuman 宣称的证伪实验。

我验证 AlphaGo 系列有没有 generic human 的围棋智力的实验并不复杂。围棋下得不好的我也可以做实验，只要 Deepmind 同意在实验上合作。目的是找出人工智能技术普遍而严重的问题。我的实验的更大价值在于较准确地测试 AlphaGo 系列的围棋智力上限，这需要一流人类棋手和 AlphaGo 系列对弈，要有使其暴露弱点的方法。而在普通的围棋比赛中，人类一流棋手是否仍有可能击败 AlphaGo 系列，也需要进一步验证。

我设计的实验，实验结果和结果分析全部公开。别人可以对实验过程和结果作分析、推敲。这才是真正的科学实验方法。如果人类棋手在我设计的公正实验条件下也能胜过 AlphaGo 系列，意味着 AlphaGo 系列所依赖的人工智能技术有严重的缺陷 [7]。这对脑和智力的研究、以及人工智能的发展都有很大价值。

我设计的实验，基于 Turing Machine 和 deep-learning 的局限性，不需要 AlphaGo 系列论文和棋谱中的信息，所以可以验证人工智能和计算机本身的局限性。但出于对科学和公众负责，论文作者应该公布跟论文有关的全面准确的实验结果：包括 AlphaGo Zero 对 AlphaGo Master，和最后版（或稳定版）的 AlphaGo Zero 对 AlphaGo Zero 的更多棋谱，等等。

基于数据的人工智能或统计方法，无法解析未来的模式演变。如果模式演变不稳定，会产生严重误导。有些以前几乎不出现的问题，模式演变后，甚至可能成为大概率事件。要了解智能的模式演变是否稳定，需要研究智力的基础理论 [8]。

三、数学危机和智力基础理论研究

科学中虽然大量使用数学，但哥德尔定理实际上已经说明数学不能用来判断科学上的正确性。Turing Machine 及其各种变种，都是数学的子集，当然也不能判断科学上的正确性。不了解 Turing Machine 上的各种问题，夸耀 Universal approximation 只是误导。

科学起源于哲学。科学的基础：Socratic method 和科学实验，都需要大量使用自然语言来描述、分析和讨论。科学论文也大量使用自然语言和哲学性思考和分析。忽视、排斥科学哲学的人，很可能用一种错误的哲学来看待科学，比如经验哲学、或实用主义。

第一次数学危机，其实跟哲学有关。科学哲学在 Karl Popper 以后，基本停滞不前。所以需要新的语言来研究智力现象以及和数学、物理学、生物学的不同和关系。对毕达哥拉斯体系和无理数的引入，以及各次数学危机的本质，也需要重新认识 [9]。

对第一次数学危机引起的智力和知识结构的演变，作重新认识，超越了数学和物理学，涉及科学的本质，是哲学和智力问题。生命需要在热力学的熵定律之上发展，这涉及生命的本质。智力也不同于能量。所以，智力科学也不同于物理学。

Neurosciences 在视觉、听觉、情绪、运动等动物和生理层次的研究上，有很多进展。但对 human specific intelligence 的研究却无法深入，也无法设计出对人工智能的证伪实验。所以智力科学也不同于普通的生物学。

人工智能的研究，从 Alan Turing，到 Geoffrey Hinton，到 Demis Hassabis，技术还在发展，但思考深度和境界却在递减。AlphaGo 系列在算法设计上有进步，在是否超越人类智力这一方向性问题的分析上，却误导了人工智能的发展。

Demis Hassabis 混淆了人类直觉和 deep-learning、reinforcement 等人工智能，宣称 AlphaGo Zero 在围棋上是 generic 的 superhuman，不知道作出科学结论前，要先接受别人的证伪实验。他只主导了一些应用的开发。而 Geoffrey Hinton 提出的是较系统的思考和方法。

Turing Machine 为计算机提供了理论基础。人工智能和计算机的局限性，可以从 Turing Machine 的局限性去研究。Geoffrey Hinton 对计算机的理解，没有达到 Turing Machine 的本质，也就无法超越 Alan Turing 的思想境界，难以理解人工智能的局限性。

要理解人工智能和计算机的局限性，研究人和人工智能更好的合作模式，需要建立更准确的智力概念和语言。

科学起源于哲学。重要科学学科的开创，均以哲学性思考为先导，慢慢引入数学。智力科学仍然基于物理学、化学、生物学的规律，也需要哲学性思考为源泉：

一、可以先研究从伽利略的 Dialogue Concerning the Two Chief World Systems [10]，到广义相对论的发展过程，去理解科学规律的本质，理解科学实验和经验测试的不同。

二、还可以研究从波义耳的 The Sceptical of Chemist，到元素周期表，到量子物理的标准模型的发展过程，去了解科学的深入和具体。

三、还可以研究从达尔文的 On the Origin of Species，到遗传学，到生物学综合的发展过程，去理解生命的本质。

智力科学，需要新的语言概念和数学形式，来理解 human specific intelligence；需要设计有针对性的实验，来收集少数的临界性和其它本质性信息，来形成和发展结构性理论。这跟 big data 技术刚好相反，也无法由 big data 技术来完成。BRAIN Initiative 项目对 big data 方法的过分倾斜，使真正的 human level intelligence 基础理论研究得不到资源而受排斥。

Big data 方法并非科学方法。对此不理解，智力科学的新思想就得不到机会来形成。把智力研究转化成科学，可以从可验证的 AlphaGo 系列的问题着手，逐步深入。

凡是享受现代教育、现代生活的人，都应该对希帕索斯和伽利略感恩，虽然他们最重要的研究工作，并没有给他们自己带来经济回报。有这样的人为了追求学术真相而牺牲自己的生活甚至生命，科学才能够持续发展到今天。

四、教育和科学研究中的危机

朗道势垒是苏联物理学家朗道招收学生的最低测试，以难度著称。成功通过朗道势垒本身并不是学术研究的成就。

国际数学奥林匹克（奥数）也起源于苏联的中学数学竞赛，但知识层次较低，针对中学的学生。奥数获奖，虽然是荣誉，但也不是学术成就。象中国的微雕艺术，是有趣而令人尊敬的技艺，但不是学术研究的主要目的。

关注点从朗道势垒到奥数，体现了所关注的知识层次的降低（也是学术境界的降低），和参与人数的进一步扩大。

作为爱好者的活动，奥数可以在年轻学生中提升、培养对数学的兴趣。但把奥数竞赛当作教育的主流指标，却有很多弊端。如果因此过分注重技巧的训练，忽视和排斥对哲学和科学方法的教育和发展，更会破坏教育的全面发展，并危及未来的科学研究 [11]。

科学研究中所需要的一些本质思考和实验方法，从奥数和其他类似的高难度、浅层次、纯技巧训练中是得不到的。

科学哲学在 Karl Popper 以后，基本停滞不前，不但导致智力研究在理论上误入歧途，也使科学研究在整体结构上越来越难以突破。很多耗资巨大的实验结束后，既证明不了什么本质，也否定不了什么本质，使新的科学理论无法成熟、诡辩又开始在学术界流行。

设计好的实验，是要找到实验上的临界点，依靠实验结果就可以建立新的理论，否定其他的可能。但在物理学的很多重要方向上，现在要设计出这样的好实验越来越难了。

遗传工程是最近几十年才发展起来的，仍充满活力。但生物基因的新研究方法，有很多也面临准确性、稳定性越来越差的情况。

即使生物基因工程能继续繁荣，能不能弥补电子和计算机产业饱和所带来的衰退？靠生物基因工程一枝独秀，新的农业社会形态会不会取代现在的后工业社会？更理想的社会形态，是能多方面发展的。

继物理学（各种机械相关产业等等）、化学（能源、材料、电子产业等等）、生物学（农业和基因工程等等），下一个关注的焦点就是：智力科学能否形成及其潜力。

但科学新领域的形成和发展，需要新的思想。Turing Test 误导了智力的理论发展。这种主观、肤浅的测试思维，也使得现在的人工智能，仍停留在经验的方法上。对未来的模式演变无法解析，而处于盲目无知的状态。

主流教科书 AI: A Modern Approach 第三版在哲学基础和测试理论上的错误，是现在人工智能研究普遍面临的问题，也影响到无人汽车的问题。不加限制说明的 driverless Car，就是 SAE level 5，这需要 human specific intelligence 来处理未来模式的演变。只有深入的基础研究才能设计证伪实验。

现在有人在研究这些基础理论吗？Neurosciences 能提供答案吗？这些问题，都源自目前教育和研究中的急功近利，及知识层次和学术境界的急剧下降。也使未来蒙上阴影。在智能挑战的时代，对逻辑和诡辩、图灵机和数学公理系统、科学哲学、和 human specific intelligence 的研究却严重不足，大学也缺乏这方面的课程和教育，是产生 Technological Singularity 误导的根源之一 [12]。

不理解语言、哲学、数学、科学的发展，就无法真正理解人类智力及其潜力。

五、数学和科学的演变

从毕达哥拉斯到伽利略，是人类文明走上可持续发展的关键。

毕达哥拉斯的抽象化思维，创立了数学。他对算术和数的定义和体系，按现代抽象代数的概念也是成立的，但不适用于几何和其他一些系统。毕达哥拉斯算术系统本身在科学上是错的，但在数学上又是可以成立的、有价值的、可以继续发展的。所以数学不同于科学，也不能简单用数学来验证科学上的对错。

希帕索斯的发现，本来也是对数学和毕达哥拉斯体系的发展。但希帕索斯事件之后，毕达哥拉斯学派本身就没有太多成就了。而其他群体能继续他们的工作，后来居上。

欧几里德几何首次创建了可以完备化的数学理论体系，并带动了阿基米德等等一大批泛希腊化时代的学术研究。

毕达哥拉斯和欧几里德两千年之后，有牛顿和莱布尼茨创建微积分，罗巴切夫斯基等人推翻欧几里德几何的唯一性，建立非欧几何 [13]。

朗道实际上把自己当作第四流的科学家：牛顿是第一流，爱因斯坦是第二流，创建量子物理的才俊们是第三流，和朗道同时代的美国物理学家理查德·費曼也是第三流。也许朗道是自谦，但他肯定无法和爱因斯坦相比。而俄罗斯历史上是出过门捷列夫的。

不知朗道该如何评价伽利略、达尔文。

按照我对学术成就的评价，创建重要学科、和重要学科中的理论体系，是第一流的工作；在重要学科中提出革命性理论体系，是第二流的工作；对主要理论体系的重要延伸或修补、或开创较小的学术分支，可列入第三流的成就。第一、第二流的学术成就非常罕见，第三流的工作已经非常了不起了。这个标准跟朗道的评价高度吻合。

物理学中，伽利略和牛顿的贡献，跟数学中的毕达哥拉斯和欧几里德相仿，属第一流。相对论和量子力学是重要学科的革命性理论体系，跟数学中的微积分和非欧几何相仿。所以，爱因斯坦的贡献是第二流的，而创建量子力学有许多人的贡献，他们的贡献只能算第三流的。量子电动力学只是小分支，所以理查德·費曼的工作也是第三流的。

牛顿和莱布尼茨在数学中的贡献是第二流的。但非欧几何的创建，由多人逐步完成，有许多人的贡献。罗巴切夫斯基和鲍耶对推翻欧几里德几何的唯一性最重要，他们和希帕索斯（发现无理数）、亚里斯多德（逻辑学）、帕斯卡（概率论）、哥德尔（哥德尔定理）相仿，仅略逊于牛顿和莱布尼茨。

拉格日朗（数学分析的成熟化）、图灵（图灵机作为计算模型）的贡献是第三流的；数学中的高斯，和物理学中的朗道，能不能挤进第三流，我不知道；祖冲之的贡献，未必能超过高斯。

数学是非常古老的学科，现在看起来很重要的一些成就，跟数学历史长河中的很多大师比起来，不一定更重要。十七、十八世纪的数学，和二十世纪前半期的物理学类似，已经十分兴旺，但基础是更早的人奠定的。

达尔文跟毕达哥拉斯和伽利略相仿。孟德尔略逊于欧几里德和牛顿，因为生物学整体理论在不久之后还需要几次重要的综合才能形成。

波义耳和门捷列夫在化学中的作用，有点类似于物理学中的伽利略和牛顿。但门捷列夫和波义耳不但时间相隔较久，而且中间有很多人的重要贡献。所以，他们

的学术贡献，略逊于伽利略和牛顿。

这里的评价，全部按照学术贡献的大小，不是对个人智力的评估。人各有特长，个人智力也不一定都表现出来，很多情况下是不可比较的。而学术贡献，是已完成的工作，可以客观分析。

整理历史上学术贡献的大小，一方面是为了研究人类智力的演变和其中重要的临界点，另一方面也是研究能促进未来发展的各种因素。现在把智力研究转化出科学的关键，在于形成准确描述智力的语言概念、哲学基础、和数学形式。

如果物理学、化学、生物学、人工智能或现有数学等等，都不能提供这些概念、基础、和形式，就需要创建新的学科。能不能提供，可以通过证伪实验来验证。

科学和数学的本质和方法，有所不同。科学兴起虽然较迟，但能弥补数学思维方法的一些严重不足。而希帕索斯和伽利略因为追求学术真相而受迫害，更应引起人们的警觉。公正的研究环境、和免受学术迫害，是科学能够持续发展的前提。相比之下，高斯虽是优秀的数学家，却缺乏追求学术真相来推动进步的求道精神，历史地位有限。

六、结语

把智力研究转化出科学的关键，在于形成准确描述智力的语言概念、哲学基础、和数学形式。科学的生命力在于设计可验证的理论。经验测试不是科学实验，经验结果不能导出科学结论。Technological Singularity 和 AlphaGo Zero 的 superhuman 宣称，可以通过严格控制条件的科学实验。

对人类智力的基础理论研究，不但可以帮助人类和人工智能更好地合作，而且可以带动脑科学、语言学、心理学、哲学、和数学等等一大批学科的发展。

在科学上观点正确却没有权势没有资源的一方，要求公正的实验是最后的生机。如果有权势资源的那方，拒绝以公正的实验来检验，等于要把正确观点逼入绝路。所以，权势和资源也应该意味着责任。

科学不是凭主观去判断该相信谁，而是靠严谨设计、严格控制条件的实验来验证。而 Technological Singularity 和 AlphaGo Zero 的 superhuman 宣称，超出了人工智能开发和围棋竞技的范围，所以需要用科学实验来检验 [14]。

~~~~~~~~~~~~

[1] AI: A Modern Approach 第三版把苏格拉底和柏拉图从哲学基础中删去，而把亚里斯多德哲学当成正确思维的基础：“Aristotle… was the first to formulate a precise set of laws governing the rational part of the mind.”(On page 5)。

Galileo Galilei 在 Dialogue Concerning the Two Chief World Systems 中已经指出了亚里斯多德哲学并不是科学的理性思维。

Immanuel Kant 对人和思维作过深入研究，他的哲学已经远远超出亚里斯多德哲学。所以哲学上的理性思维，实际上也无法用亚里斯多德哲学正确表达。

Gödel's theorems 更说明了逻辑实证主义有问题，亚里斯多德哲学甚至不能完整表达数学中的理性思维。

所以，亚里斯多德哲学在哲学、数学、科学上都不能准确表达理性思维，当然不可能是 "a precise set of laws governing the rational part of the mind"。

这个问题不澄清，不但人工智能研究受到误导，脑科学、语言学、心理学、哲学等等一大批学科的发展会受到阻碍。越来越多的领域试图用非理性行为来解释现象，实际上很多是出于对理性的误解，看不清非理性表象下面真正的理性原理。
~~~~~~~~~~~~

[2] 按政府报告 Leucotomy in England and Wales, 1942-1954，该手术的 9284 名患者中，41%恢复或有明显改善，28%稍微改善，25%显示无变化，2%恶化，4%死亡。

但这个统计明显有问题。如此破坏性的脑手术，当然会破坏 personality 和 intelligence。所以，这里的 25%显示无变化，不是指 personality 和 intelligence 未受破坏，仅仅是指为控制精神状态而特殊定义下的 clinical condition 无变化。同理，41%恢复或有明显改善和 28%稍微改善的，也只是特殊定义下的 clinical condition 有改善，personality 和 intelligence 实际上都有损害。

本来 leucotomy 手术就是为了控制精神状态，而不惜损害脑功能。

也有冠名 Renato M.E. Sabbatini 教授的文章指：Even lobotomy's preponents admitted that only one third of the operated patients would improve, while one-third remained the same, and one-third got worst 。这个数字跟 Leucotomy in England and Wales, 1942-1954 明显不同。（http://www.cerebromente.org.br/n02/historia/lobotomy.htm）

但这里的 one third would improve 和 one-third remained the same，也仅仅指特殊定义下的 clinical condition，不是指 personality 和 intelligence 未受破坏。

所以，由于缺乏正确描述 personality 和 intelligence 的语言，就无法对 leucotomy 手术和使用基因技术的 BRAIN Initiative 方法的破坏性后果作研究。这种情况急需改变。

[3] 许峰雄在 Cracking Go 文中更早提出。但他的依据很多来自开发 Deep Blue 的经验，对围棋不一定成立。我的理论分析，不依赖经验，反而能成立。我的理论分析还可以进一步延伸：AlphaGo 系列虽然赢棋，但只要不是最优解，就可能走的是无理棋；可以设计严格控制条件的严谨实验来检验 AlphaGo 系列是否有通用智力潜力，是否超人。

[4] 根据 http://www.alphago-games.com/，AlphaGo Zero 对 AlphaGo Zero 中只有一局是最后版（或稳定版）的。在 https://www.101weiqi.com/chessbook/player/38348/也有相关棋谱。

[5] AlphaGo Master 不是生命，AlphaGo Master 进化速度慢，不能说明生物的演化本质。部分原因是裁定围棋胜负的规则是由人类定义的，而生物演化的规则并非由人类定义。

[6] http://www.alphago-games.com/上，Full Strength of Alphago Zero, i.e. Final Form，被称为 40 Blocks。20 Blocks 被称为 Not Full Strength of Alphago Zero。更新结构的系统击败 Alphago Zero 也是非常可能的。

[7] 我选择围棋作实验工具，是因为可以用围棋验证人工智能和计算机的一些基础科学理论，跟军事战略和战争艺术无关。围棋是和平的竞技，高度抽象而又直观。既古老又现代，能吸引亚洲以外的电脑研究。已经成为跨文化研究、交流的平台。

第一轮实验后，开发者可能修改软件。所以有些实验证据只能在第一轮实验中收集。第一轮实验需要跟最强版本电脑围棋作较大规模的实验，以收集足够的实验证据。

我对 AlphaGo 系列的实验不是白盒测试，不涉及知识产权。如果 Google 愿意公正地检验，可以仿照美国国会图书馆先例，先向国会图书馆递交 AlphaGo 柯洁版、AlphaGo Zero、和最新 AlphaGo 版本所需要的所有硬件、软件和文档，能捐款给图书馆用于维护则更好。以后在第三方场地实验时，可以向国会图书馆核对软件、硬件和文档等版本。Human level artificial intelligence 的检验，是人类史上的里程碑。如果无人能对 AlphaGo 系列证伪，国会图书馆就有继续保留其版本的价值。其他人有实验想法，可以先做。如果其他人都失败了，就没有道理拒绝我的实验方案。

这也为美国政府了解人工智能的最新进展提供了渠道，并为以后的人工智能测试建立了一个范式。一旦相关实验设施建立好并熟悉了这种流程，以后重要的人工

智能测试费用会大大降低。

[8] 根据新闻报道，2012 年时，就有盲人使用自动驾驶汽车；2015 年时，有盲人单独使用无人汽车。根据公开的事实，已经知道当时无人驾驶汽车没有成熟，对人工智能和无人驾驶汽车有误判和误导。也说明之前的经验测试不可靠。

我以前写过:"Go gaming is strictly defined within a very small space. Industrial automations are typically designed in well controlled environments, but not strictly defined. Car driving is regulated, but the environment is not well controlled"。无人驾驶汽车中的问题，要远远比电脑围棋复杂。而即使在电脑围棋中，仍有问题可以被证伪。

另外，无人驾驶汽车的测试，完全不同于其他汽车。汽车中的动力（油动或电动）、机械、电子等设计，早就具备坚实的科学基础。而无人驾驶的人工智能，没有科学基础，也未经严谨的科学实验，危险性很大。现在使用的是经验测试，不可靠。需要新的严格控制条件的科学实验来证伪。

[9] 物理学界有人质疑时空是连续的。有人更以此质疑我的智力理论。对我的质疑毫无道理。第一，要看非连续的时空理论是否可验证。很多现代物理理论根本没法验证，没法验证的就不是科学。象围棋走法的数量超过宇宙的原子数就是这种理论，因为对宇宙原子数量的推测根本不可靠，也无法验证。第二，著名物理学家的观点不一定是对的，如牛顿迷恋炼金术。第三，时空的定义和时空的测量是两回事，我的智力理论跟时空在测量上是否是连续根本没有关系，我研究的是无理数在科学及其未来发展上是否不可缺少，这是智力科学的问题。这再次说明智力科学和物理学是不同的。

[10] Dialogue Concerning the Two Chief Word Systems 实际上把科学建立在 Socratic Method 和科学实验的基础上。文中有错误，特别是关于潮汐的。科学不是完美的，但科学需要能改进错误的方法。文中倡导的思考和方法，可以发现和纠正关于潮汐的错误。

伽利略文中的错误，和第谷的错误不一样。伽利略的思想方法是对的，理论体系也是对的。而第谷的思想方法是错的，理论体系也是错的。如果采用第谷的理论，就无法发展出牛顿的物理基本理论。混淆人工智能和人类智力，也无法发展出智力科学。

第谷的观察工作仍有价值是侥幸，因为地心说、日心说、第谷理论在观察方法上没有争议。而很多科学实验，需要严格控制实验条件。错误的实验条件会导致错误的结论。

所以该文提供了科学的良好范本。科学不应该戴上圣人或权威的面具，科学应该放下身段来平等讨论。

[11] 近二、三百年，俄罗斯的人口规模和 talent pool 的规模远超过其他欧洲国家，在人才的衡量和选拔上有优势。中国的人口规模更大，远超过世界上的其他发达国家。奥数在中国受到追捧，也不难理解。但靠人才优势以技巧争锋，也有弊端。

美国人口规模也不小，近几十年更有财力在全世界范围吸引人才。所以，最近几十年，美国在 talent pool 上的规模优势，其实比中国更强。美国以前在量子物理、天体物理、遗传学、计算机工程等需要新思想的学科上，比前苏联强。而近年则越来越偏向技巧，排斥思想，而在走下坡路。

[12] 根据哥德尔定理，图灵机和数学本身都有严重的不足。Universal approximation 并不存在。Technological Singularity 只是欲望，并无科学依据。AlphaGo Zero 的 superhuman 宣称则也可以通过严格控制条件的严谨科学实验来证伪。

[13] 高斯不但没有公布自己在非欧几何上的想法，也没有公布 1819 年交给他的 Ferdinand Karl Schweikart 的类似想法。所以高斯在非欧几何上，不但无功，反而有过。根本原因，在于高斯太注重坚实的技巧性，不了解尚未成熟的思想体系的价值，也不重视苏格拉底方法对学术发展的作用。高斯不但缺乏学术上的眼光，也达不到罗巴切夫斯基的思想境界和层次。

罗巴切夫斯基和鲍耶推翻欧几里德几何的唯一性，已是 1830 年前后了；他们在非欧几何上贡献最大。黎曼几何在物理相对论研究中，有价值。但黎曼在非欧几何中，只是作延伸性工作，贡献次于罗巴切夫斯基和鲍耶。贝尔特拉米为非欧几何带来新的语义和完整性，也很重要。

[14] 人工智能最大的危险，不是其太强大可能控制或消灭人类，而是大规模应用不成熟的人工智能可能带来巨大灾难。所以，能通过严谨设计的电脑围棋科学实验来研究智力基础理论问题，对人类是一种幸运。